

Publishing Dataset to GGBN Protocol

Step 1. Downloading the dataset

Tissue and associated specimen data is stored within the Research Department Database.

Navigate to <https://research.botanicgardens.org> and log in.

Once logged in, navigate to <https://research.botanicgardens.org/admin/query.php> (just add query.php to end of home page web address)

Paste the query from the box below into the web page's box and select "submit"

```
SELECT CONCAT("Tissue ", wd.work_id, "-", wd.work_individual_data_id) AS "catalogNumber", wid.UUID AS "materialSampleID", "Tissue and DNA Bank" AS "collectionCode", "DBG" AS "institutionCode",
"MaterialSample" AS "basisOfRecord", wid.collection_type AS "materialSampleType", dprep.preparationType AS "preparationType", tw.preservation AS "preservationType", pmts.permit_availability AS
"permitStatus", pmmt.permit_type AS "permitType", "For full permit text, please contact Denver Botanic Gardens Research and Conservation directly" AS "permitText", vw.collector_number AS
"recordNumber", u.fullname AS "recordedBy", wd.date_collected AS "eventDate", cntryiso.IsoCode AS "countryCode", cntryiso.Country AS "country", wd.latitude AS "decimalLatitude", sts.state AS
"stateProvince", cnts.county AS "county", wd.longitude AS "decimalLongitude", wd.elevation AS "minimumElevationInMeters", f.family AS "family", sl.synonym_species_name AS "scientificName",
SUBSTRING_INDEX(sl.synonym_species_name, ' ', 1) AS "genus", SUBSTRING_INDEX(SUBSTRING_INDEX(sl.synonym_species_name, ' ', 2), ' ', -1) AS "specificEpithet", "species" AS "taxonRank",
vw.KHD_accession_number AS "catalogNumberVoucher", vw.voucher_GUID AS "occurrenceID", wd.specific_locality AS "locality", "same population" AS "relationshipOfResource",
CONCAT("catalogNumber=", vw.KHD_accession_number, "&institutionCode=KHD&collectionCode=vascular
plants&accessPoint=http://ipt.vertnet.org:8080/ipt/archive.do?r=khd_vascular&guid=", vw.voucher_GUID) AS "resourceRelationshipID", wid.UUID AS "resourceID", vw.voucher_GUID AS "relatedResourceID",
"tissue from many individuals collected from same population as individual taken as voucher specimen" AS "relationshipRemarks", "Permit is available" AS "permitStatusQualifier", wd.date_collected AS
"preparationDate", acsp.redact_locality

FROM work_done AS wd
LEFT JOIN work_individuals AS wi ON wd.work_id = wi.work_id
LEFT JOIN work_individual_data AS wid ON wi.work_individual_id = wid.work_individual_id
LEFT JOIN tissue_work AS tw ON wd.work_id = tw.work_id
LEFT JOIN agencies AS agen ON wd.agency_id = agen.agency_id
LEFT JOIN dna_preparations AS dprep ON tw.preparation_type = dprep.preparation_id
LEFT JOIN permits AS pmts ON wd.permit_id = pmts.permit_id
LEFT JOIN permit_types AS pmmt ON pmts.permit_type_id = pmmt.permit_type_id
LEFT JOIN countryiso AS cntryiso ON wd.ID_Country = cntryiso.ID_Country
LEFT JOIN states AS sts ON wd.state_id = sts.state_id
LEFT JOIN counties AS cnts ON wd.county_id = cnts.county_id
LEFT JOIN voucher_work AS vw ON wd.work_id = vw.work_id
LEFT JOIN users AS u ON vw.primary_collector = u.user_id
LEFT JOIN work_landowners AS wldo ON wd.work_id = wldo.work_id
LEFT JOIN landowners AS lndo ON wldo.landowner_id = lndo.landowner_id
LEFT JOIN accepted_species AS acsp ON wd.accepted_species_id = acsp.accepted_species_id
LEFT JOIN families AS f ON acsp.family_id = f.family_id
LEFT JOIN synonym_list AS sl ON wd.accepted_species_id = sl.accepted_species_id

WHERE NOT wid.collection_type = "herbarium" AND NOT wid.collection_type = "seed" AND NOT wid.collection_type = "dna" AND wid.collection_type IS NOT NULL AND wid.material_remaining = 1 AND
vw.work_id IS NOT NULL
GROUP BY wid.UUID
;
```

On the results page, select "Download as CSV"

Locate and open the .csv file that was downloaded.

Step 2. Checking and modifying the dataset

In order to continue, herbarium specimens must be curated to the point that they have been assigned a catalog number (barcode). If catalogNumberVoucher is blank, these samples are not ready to be published on GGBN.

Check each column for empty records or missing data. If data is missing, contact database associate. In this case, it is likely that default values or pieces of data were not populated during the processing of samples and specimens.

Fields that are commonly left blank, but should be populated within the database (not just in the spreadsheet) include:

- permitStatus
- permitType
- countryCode
- country
- locality
- minimumElevationInMeters

These fields are not included in the initial data entry stages and require secondary action. Contact database associate to do so.

IMPORTANT The last column is called `redact_locality`. Sort the dataset by this column, from greatest to lowest value. For all records that have the value "1" recorded in `redact_locality`, delete the decimalLatitude and decimalLongitude before publication.

Save the file here: Q:\Research\Data_Management\GGBN\Published to GGBN

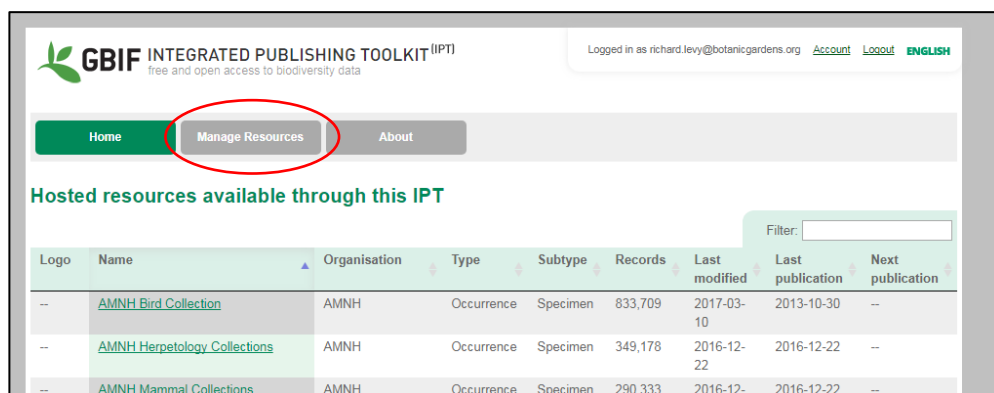
Use the naming convention `YYYYMMDD_dataset_DBG Tissue And DNABank_GGBN`

Step 3. Uploading the file

Navigate to VertNet's IPT instance and log in. Users will need to contact VertNet for access.

<http://ipt.vertnet.org:8080/ipt/>

Select "Manage Resources"



GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)
free and open access to biodiversity data

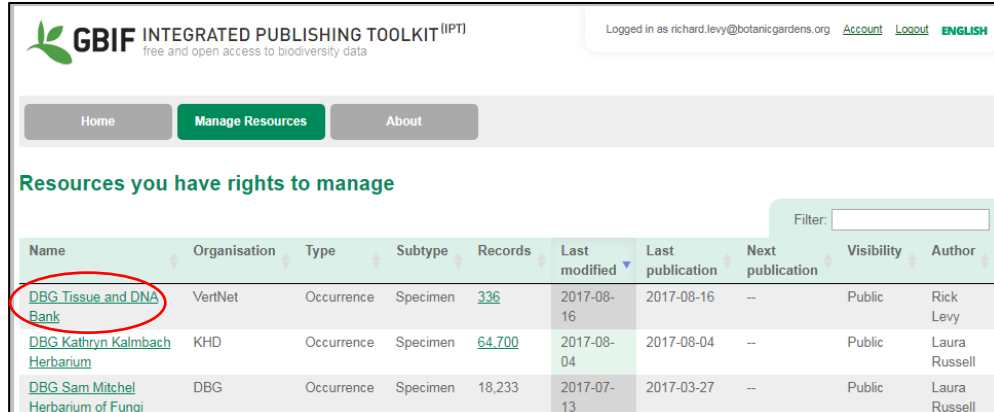
Logged in as richard.levy@botanicgardens.org [Account](#) [Logout](#) [ENGLISH](#)

Home **Manage Resources** About

Hosted resources available through this IPT

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
--	AMNH Bird Collection	AMNH	Occurrence	Specimen	833,709	2017-03-10	2013-10-30	--
--	AMNH Herpetology Collections	AMNH	Occurrence	Specimen	349,178	2016-12-22	2016-12-22	--
--	AMNH Mammal Collections	AMNH	Occurrence	Specimen	290,333	2016-12-	2016-12-22	--

Select "DBG Tissue and DNA Bank"



GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)
free and open access to biodiversity data

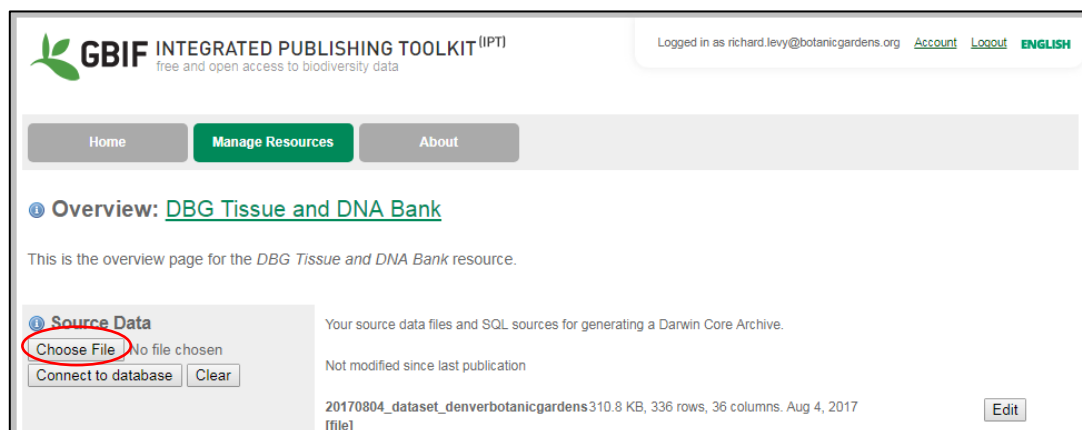
Logged in as richard.levy@botanicgardens.org [Account](#) [Logout](#) [ENGLISH](#)

Home **Manage Resources** About

Resources you have rights to manage

Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication	Visibility	Author
DBG Tissue and DNA Bank	VertNet	Occurrence	Specimen	336	2017-08-16	2017-08-16	--	Public	Rick Levy
DBG Kathryn Kalmbach Herbarium	KHD	Occurrence	Specimen	64,700	2017-08-04	2017-08-04	--	Public	Laura Russell
DBG Sam Mitchell Herbarium of Fungi	DBG	Occurrence	Specimen	18,233	2017-07-13	2017-03-27	--	Public	Laura Russell

Under "Source Data" select "Choose File". Navigate to the new file for upload.



GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)
free and open access to biodiversity data

Logged in as richard.levy@botanicgardens.org [Account](#) [Logout](#) [ENGLISH](#)

Home **Manage Resources** About

Overview: [DBG Tissue and DNA Bank](#)

This is the overview page for the *DBG Tissue and DNA Bank* resource.

Source Data

Choose File No file chosen

[Connect to database](#) [Clear](#)

Your source data files and SQL sources for generating a Darwin Core Archive.

Not modified since last publication

20170804_dataset_denverbotanicgardens310.8 KB, 336 rows, 36 columns. Aug 4, 2017 [file] [Edit](#)

Once the Source File has been selected, select “Add”.

The screenshot shows the GBIF Integrated Publishing Toolkit (IPT) interface. At the top, there's a header with the GBIF logo and the text "free and open access to biodiversity data". A navigation bar includes "Home", "Manage Resources", and "About". The main content area is titled "Overview: DBG Tissue and DNA Bank" and includes a sub-section "Source Data". In the "Source Data" section, there's a "Choose File" button, a text input field containing "20170804_d..._GGBN.csv", and an "Add" button which is circled in red. To the right of the "Add" button is a "Clear" button. Further right, there's text indicating the file was "Not modified since last publication" and a summary of the dataset: "20170804_dataset_denverbotanicgardens 310.8 KB, 336 rows, 36 columns. Aug 4, 2017 [file]". An "Edit" button is located at the bottom right of the "Source Data" section.

After clicking “add” the user is brought to the Edit Source Data Format page.

Change the “Date Format” to be “MM/DD/YYYY”. Then select “Save”.

The screenshot shows the "Edit Source Data Format" page in the GBIF IPT. At the top, there's a green success message: "Added new file source >>20170804_dataset_dbgtissueanddnabank_ggbn<<.". Below this, the "Resource Title" is "DBG Tissue and DNA Bank". The "Source Data" section has a sub-header "Edit your source data format". The "Source Name" field contains "20170804_dataset_dbgtissueanddnabank_ggbn". To the right, a table of metadata is displayed: "Readable" (checked), "Columns" (36), "File" (/home/iptdata/resources/tissue_and_dna_bank/sources/20170804_dataset_dbgtissueanddnabank_ggbn.txt), "Size" (310.8 KB), "Rows" (336), "Modified" (Aug 22, 2017 11:52:56 AM), and "Source log" (Download). Below the metadata is an "Analyze" button. The "Number of Header Rows" is set to "1". The "Field Quotes" field is empty. The "Character Encoding" is set to "UTF-8". The "Field Delimiter" is set to ".". The "Multi-value Delimiter" is empty. The "Date Format" is set to "MM/DD/YYYY", which is circled in red. At the bottom, there are three buttons: "Save" (circled in red), "Cancel", and "Delete source file".

Step 4. Mapping the Fields

In addition to the Darwin Core Occurrence, 5 Extensions should be added for mapping. They can be selected and added from the dropdown list on the left side of the page under “Darwin Core Mappings”.

Darwin Core Mappings

Your mapping between the source data and Darwin Core terms.

Not modified since last publication

Core

Darwin Core Occurrence

22 terms mapped to 20170804_dataset_denverbotanicgardens. Aug 7, 2017

Extensions

GGBN Preservation Extension

1 terms mapped to 20170804_dataset_denverbotanicgardens. Aug 7, 2017

GGBN Permit Extension

4 terms mapped to 20170804_dataset_denverbotanicgardens. Aug 7, 2017

GGBN Preparation Extension

2 terms mapped to 20170804_dataset_denverbotanicgardens. Aug 7, 2017

GGBN Material Sample Extension

1 terms mapped to 20170804_dataset_denverbotanicgardens. Aug 7, 2017

Darwin Core Resource Relationship

5 terms mapped to 20170804_dataset_denverbotanicgardens. Aug 7, 2017

For each set of mappings, the user must manually map the fields from the original dataset to the accepted term. Select the “Edit” button to the right of each mapping to do so.

Map the following fields

Darwin Core Occurrence:

occurrenceID	→	materialSampleID
institutionCode	→	institutionCode
collectionCode	→	collectionCode
basisOfRecord	→	basisOfRecord
materialSampleID	→	materialSampleID
catalogNumber	→	catalogNumber
recordNumber	→	recordNumber
recordedBy	→	recordedBy
eventDate	→	eventDate
country	→	country
countryCode	→	countryCode
stateProvince	→	stateProvince
county	→	county
locality	→	locality
minimumElevationInMeter	→	minimumElevationInMeters
decimalLatitude	→	decimalLatitude
decimalLongitude	→	decimalLongitude
scientificName	→	scientificName
family	→	family
genus	→	genus
specificEpithet	→	specificEpithet
taxonRank	→	taxonRank

GGBN Preservation Extension

occurrenceID	→	materialSampleID
preservationType	→	preservation

GGBN Permit Extension

occurrenceID	→	materialSampleID
permitType	→	permitType
permitStatus	→	permitStatus
permitStatusQualifier	→	permitStatusQualifier
permitText	→	permitText

GGBN Preparation Extension

occurrenceID	→	materialSampleID
preparationType	→	preparationType
preparationDate	→	preparationDate

GGBN Material Sample Extension

occurrenceID	→	materialSampleID
--------------	---	------------------

Darwin Core Resource Relationship

occurrenceID	→	materialSampleID
resourceRelationshipID	→	resourceRelationshipID
relatedResourceID	→	relatedResourceID
relationshipOfResource	→	relationshipOfResource
relationshipRemarks	→	relationshipRemarks
scientificName	→	scientificName

Step 5. Preview and Publish the Pending Version

Preview the pending version by selecting “Preview”. Verify that everything looks correct.

Once previewed and verified, select “Publish”.

The screenshot shows a user interface for managing resource versions. On the left, under the 'Published Versions' section, there is a 'Publish' button circled in red. Below it is an 'Auto-publishing' section with a dropdown menu set to 'Select interval'. To the right, there is a table comparing the 'Current version' (1.5) with the 'Pending version' (1.6). The 'Current version' row has buttons for 'View' (with a green checkmark), 'Download', and 'Show'. The 'Pending version' row has a 'Preview' button circled in red. Above the table, there is a description: 'A preview of your pending published version compared with the current version if existing.'

	Current version	Pending version
Version	1.5 View ✓	1.6 Preview
Published on	Aug 16, 2017	-
Publication log	Download	-
Publication report	Show	-

Once the user selects “Publish” a window opens a prompt. In the window briefly summarize what has changed in the new version to be published. Then select “Yes” when you are ready to publish. On the next page that opens, the IPT will run through the publishing process and notify the user when it has finished successfully.

The dialog box is titled "Are you sure?" and contains the following text:

You are about to publish a new **minor** version of this resource. If there have been scientifically significant changes to the resource since the last publication, you should cancel, assign this resource a new DOI, and then publish again.

Before you publish, you can assist researchers using your data by summarizing what has changed in this version.

Do you still want to publish?

Please summarize what has changed in this version. You can always edit this later.

At the bottom, there are two buttons: "Yes" and "No".